

## Mục lục

MỤC LỤC . . . . .	
DANH SÁCH HÌNH VẼ . . . . .	
DANH SÁCH BẢNG BIỂU . . . . .	
DANH MỤC KÝ HIỆU VÀ VIẾT TẮT . . . . .	
<b>CHƯƠNG 1    CƠ SỞ LÝ THUYẾT</b>	<b>1</b>
1.1   Các phương pháp khai thác nhân tạo . . . . .	1
1.1.1   Bơm điện li tâm chìm . . . . .	2
1.1.1.1   Ứng dụng . . . . .	2
1.1.1.2   Ưu điểm và hạn chế . . . . .	2
1.1.2   Gas lift . . . . .	3
1.1.2.1   Ứng dụng . . . . .	5
1.1.2.2   Ưu điểm và hạn chế . . . . .	6
1.1.3   Một số phương pháp khác . . . . .	7
1.2   Trí tuệ nhân tạo . . . . .	8
1.2.1   Giải thuật di truyền . . . . .	8
1.2.1.1   Nguyên lý giải thuật . . . . .	8
1.2.1.2   Thuật toán di truyền . . . . .	10

1.2.2	Bài toán thời gian . . . . .	13
1.2.3	Mạng neuron truyền thẳng đa lớp . . . . .	15
1.2.3.1	Cấu trúc cơ bản . . . . .	16
1.2.3.2	Hàm kích hoạt . . . . .	17
1.2.3.3	Hàm tối ưu . . . . .	19
1.2.3.4	Phương pháp đánh giá . . . . .	20
1.3	Đường cong đặc tính gas lift . . . . .	20
1.3.1	Đường cong đặc tính và bài toán tối ưu . . . . .	21
1.3.2	Mô hình toán học . . . . .	22
1.3.3	Giải pháp số . . . . .	26
1.4	Thuật toán tối ưu . . . . .	27

## Danh sách hình vẽ

1.1	Lưu lượng khai thác tối đa theo độ sâu . . . . .	2
1.2	Hệ thống bơm ép gas lift (bề mặt) . . . . .	4
1.3	Sơ đồ giếng khai thác với gas lift . . . . .	5
1.4	Thuật toán di truyền . . . . .	12
1.5	Cấu trúc của một mạng truyền thẳng . . . . .	16
1.6	Hàm <i>Tanh</i> và <i>Sigmoid</i> . . . . .	18
1.7	Hàm ReLU . . . . .	19
1.8	Đường đặc tính gas lift . . . . .	22

**Danh sách bảng**

## Danh mục ký hiệu và viết tắt

$\bar{\rho}$      Tỷ trọng hỗn hợp

$f$      Hệ số ma sát

$P_r$      Áp suất vỉa

$P_{wf}$      Áp suất dòng vào tại đáy giếng

$q_g$      Lưu lượng khí bơm ép (MSCF/d)

$q_l$      Lưu lượng khai thác (lỏng) (STB/d)

$q_o$      Lưu lượng dầu (STB/d)

$Q_{gav}$      Tổng lượng khí có sẵn

$u_m$      Vận tốc hỗn hợp (ft/s)

AI     Artificial Intelligence

ANN     mạng thần kinh nhân tạo

D     Đường kính tubing (in)

ESP     Electrical Submersible Pump

GA     Giải thuật di truyền

GLPC     Gas Lift Performance Curve

J     Chỉ số năng suất (STB/d/psi)

P     Áp suất

WC     Hàm lượng nước

WOR Tỉ số nước dầu

z Chiều sâu giếng khoan

# CHƯƠNG 1. CƠ SỞ LÝ THUYẾT

## 1.1. Các phương pháp khai thác nhân tạo

Thông thường, trong giai đoạn đầu của một giếng khai thác dòng lưu chất từ vỉa sẽ khai thác nhờ áp suất vỉa (nếu áp suất vỉa lớn hơn tổng áp suất mất mát của hệ thống khai thác), giai đoạn này được gọi là khai thác tự phun. Nếu như áp suất vỉa không còn đủ lớn để có thể thắng được áp suất mất mát, dòng chảy trong giếng sẽ “chết” và giếng không thể tiếp tục khai thác tự phun được nữa. Hai nguyên nhân chính dẫn đến dòng trong giếng bắt đầu “chết” chính là:

- Áp suất dòng vào ở đáy giếng giảm xuống thấp hơn tổng áp suất mất mát của giếng,
- Tổng áp suất mất mát của giếng lớn hơn áp suất đáy giếng nhỏ nhất để có thể đưa được dòng lưu chất lên bề mặt.

Trường hợp đầu tiên xảy ra khi lưu chất được khai thác khỏi vỉa sẽ dẫn tới áp suất của vỉa suy giảm theo thời gian cho tới khi không còn đủ lớn để thắng được áp suất mất mát của hệ thống khai thác. Đối với trường hợp thứ hai, thường do tăng mức độ cản trở dòng chảy trong hệ thống, có hai nguyên nhân dẫn đến điều này:

- Tỷ trọng của dòng lưu chất tăng do lượng khí đồng hành giảm,
- Các vấn đề cơ học như kích thước tubing nhỏ, lắng đọng cặn ...

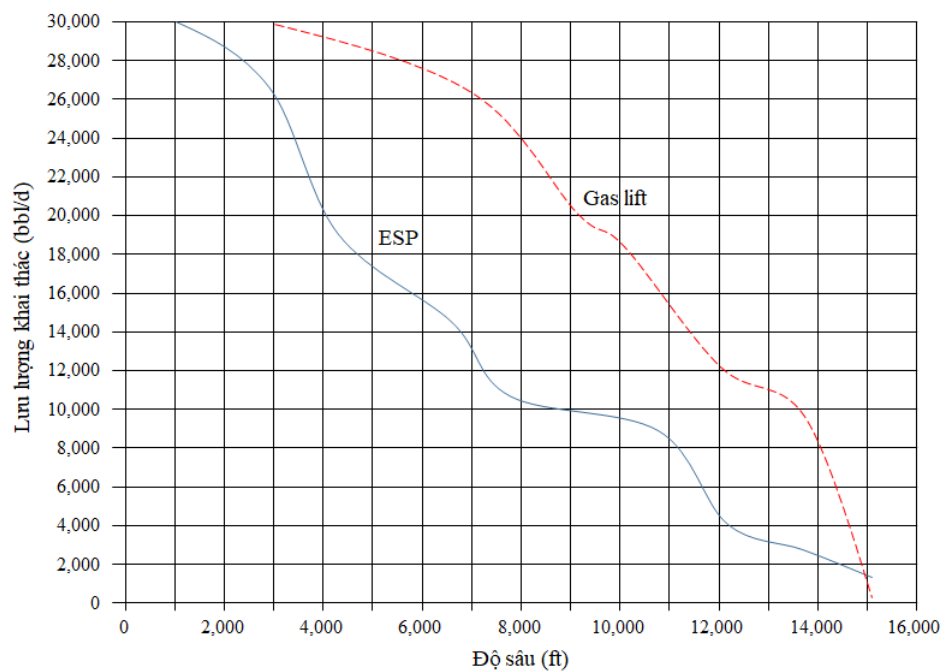
Khai thác nhân tạo được thực hiện khi dòng chảy trong giếng bắt đầu “chết” hoặc tăng lưu lượng khai thác của giếng. Hai phương pháp phổ biến nhất thường được lựa chọn để thực hiện khai thác nhân tạo là bơm điện li tâm chìm (ESP); một máy bơm sẽ được đặt xuống lòng giếng dưới mực lưu chất trong giếng, bơm tăng áp suất để thắng được tổn thất áp suất trong hệ thống; và bơm ép khí từ bề mặt vào trong tubing khai thác để dòng lưu chất dễ dàng phun lên bề mặt hơn (gas lift). Ngoài ra vẫn còn một số phương pháp khai thác nhân tạo khác nhưng không được sử dụng phổ biến.

### 1.1.1. Bơm điện li tâm chìm

#### 1.1.1.1. Ứng dụng

Được đưa vào ứng dụng trong thực tiễn từ năm 1920 tại mỏ Oklahoma, ESP có thể nâng lưu lượng khai thác thêm tới 1000 thùng trên ngày, gấp khoảng 2 đến 3 lần bơm cần (rod pump).

Ngày nay, bơm điện li tâm chìm thường được ứng dụng trong bơm ép nước (trên bờ), khai thác trên biển hoặc trong những trường hợp có sẵn nguồn điện và khai thác với lưu lượng lớn. ESP có thể được lắp đặt ở độ sâu 1,000 ft đến 10,000 ft đồng thời tăng lưu lượng khai thác thêm từ 200 bbl/d đến 20,000 bbl/d (Hình 1.1). Khả năng độ sâu và lưu lượng có thể sử dụng ESP là 15,000 ft và 30,000 bbl/d.



Hình 1.1. Lưu lượng khai thác tối đa theo độ sâu

#### 1.1.1.2. Ưu điểm và hạn chế

Một vài ưu điểm của bơm điện li tâm chìm có thể kể đến như sau:



- Thích hợp với khai thác lưu lượng lớn,
- Hiệu suất cao,
- Sử dụng tốt trong giếng khoan định hướng,
- Có thể sử dụng trong điều kiện khu dân cư,
- Thích hợp với hoạt động khai thác trên biển,
- Dễ dàng xử lý ăn mòn và lắng đọng.

ESP cũng có một vài hạn chế:

- Phụ thuộc vào nguồn cung cấp năng lượng,
- Cần được vận hành trong điều kiện ổn định,
- Dễ bị ảnh hưởng bởi cát hay các vật chất gây ăn mòn,
- Quá trình sửa chữa khó khăn,
- Bị hạn chế trong vùng có nhiệt độ cao,
- Hiệu suất thấp đối với dầu có độ nhớt cao,
- Chi phí lắp đặt cao.

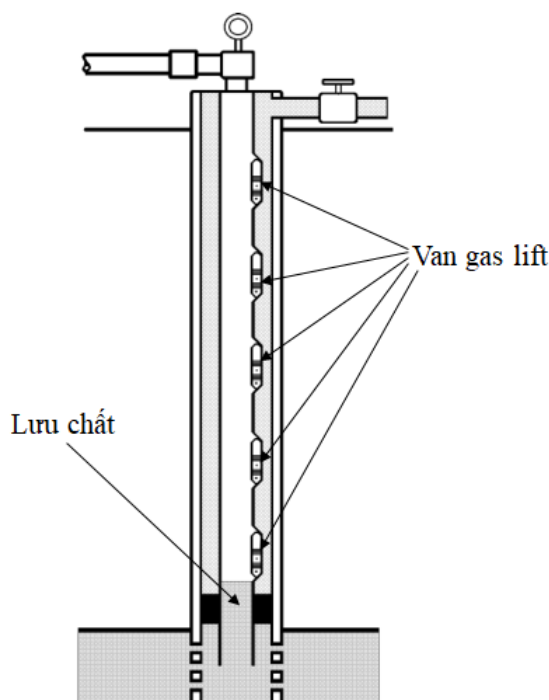
### **1.1.2. Gas lift**

Gas lift là phương pháp bơm ép khí vào cột lưu chất trong giếng để lưu chất có thể dễ dàng được khai thác hơn. Hệ thống bề mặt cơ bản cho phương pháp gas lift được thể hiện như trong Hình 1.2. Hầu hết khí được sử dụng trong bơm ép là khí tự nhiên có đặc tính trơ, được bơm ép vào trong tubing tại những vị trí đã được thiết kế trước, dựa trên ba nguyên lý chủ yếu: giảm tỉ trọng hỗn hợp lưu chất, năng lượng giãn nở của khí và thay thế vị trí. Từ đó dòng lưu chất tại đáy giếng có thể tiếp tục được đưa lên bề mặt và giếng bắt đầu được khai thác trở lại hoặc nâng cao năng suất khai thác. Có thể chia gas lift thành hai loại là gas lift định kỳ và gas lift liên tục.

Đối với gas lift liên tục, dòng khí được bơm liên tục có kiểm soát vào trong tubing trong khi gas lift định kỳ sẽ thực hiện bơm khí theo chu kỳ nhất định khi lưu chất trong tubing có



mỗi giếng dưới điều kiện tổng lượng khí bơm ép xác định để đạt được lưu lượng dầu tối đa trong điều kiện bình tách có thể chịu được.



Hình 1.3. Sơ đồ giếng khai thác với gas lift

#### 1.1.2.1. Ứng dụng

Lần đầu tiên gas lift được đưa vào trong khai thác năm 1846 tại Mỹ, tuy nhiên để sử dụng khai thác nước là chủ yếu. Đến năm 1930 gas lift mới thực sự được đưa vào ứng dụng trong ngành công nghiệp dầu khí cùng với sự ra đời của các mẫu van gas lift khác nhau. Vào những năm này gas lift được sử dụng do hai nguyên chính:

- Tăng mạnh lưu lượng khai thác trong khi năng lượng vỉa bắt đầu suy kiệt,
- Đem lại nhiều lợi ích kinh tế khi lượng khí bơm ép sau sử dụng có thể được tách và sử dụng lại, ít bị thất thoát trong quá trình sử dụng.

Trong ngành công nghiệp dầu khí hiện đại, gas lift có những ứng dụng chính như sau:

- Đưa giếng trở lại khai thác sau khi đã kết thúc giai đoạn khai thác tự nhiên,
- Tăng lưu lượng khai thác,

- Loại bỏ thành phần lỏng trong các giếng khai thác khí,
- Xử lý những giếng lắng đọng cát trong giai đoạn bắn mở vỉa,
- Khai thác nước cho quá trình bơm ép nước.

#### 1.1.2.2. Ưu điểm và hạn chế

Gas lift được xem là một trong những phương pháp phổ biến nhất trong các phương pháp khai thác nhân tạo, phù hợp cho những giếng có độ sâu lớn mà những phương pháp khác không thể thực hiện được. Một số ưu điểm chính có thể kể đến như sau:

Gas lift liên tục:

- Tận dụng tối đa lợi ích do nguồn khí đem lại,
- Khai thác lưu lượng lớn,
- Thiết bị phổ biến,
- Sử dụng được cho giếng có nhiều cát,
- Có thể thu hồi van bằng tubing hay dây cáp.

Gas lift định kì:

- Sử dụng được ở điều kiện áp suất dòng vào thấp hơn so với gas lift liên tục,
- Thiết bị phổ biến,
- Có thể thu hồi van bằng tubing hay dây cáp.

Đồng thời gas lift cũng mang một vài hạn chế như:

Gas lift liên tục:

- Cần nguồn cung cấp khí ổn định,
- Không thể triển khai đối với giếng có áp suất quá thấp hoặc có thể nhưng lưu lượng khai thác không khả quan.

Gas lift định kì:

- Không thể cho dòng khai thác lưu lượng lớn,
- Có thể gây ra hiệu ứng vọt xung trên bề mặt (surge),
- Cần nguồn cung cấp khí ổn định.

Ngoài ra, chi phí ban đầu để triển khai thực hiện gas lift khá cao, quá trình bảo trì và vận hành không dễ dàng.

### ***1.1.3. Một số phương pháp khác***

Một số phương pháp khai thác nhân tạo khác có thể kể đến như bơm cần, bơm thủy lực hay bơm trục vít. Tuy nhiên, hiện tại các phương pháp không còn được sử dụng phổ biến như hai phương pháp bơm điện li tâm chìm và gas lift. Ưu, nhược điểm của các phương pháp này có thể được kể đến như sau:

Bơm cần (sucker-rod pump)

Ưu điểm	Hạn chế
<ul style="list-style-type: none"> <li>• Hiệu suất khá cao</li> <li>• Chi phí thấp</li> <li>• Giảm thiểu ăn mòn</li> <li>• Linh hoạt</li> <li>• Dễ dàng triển khai cứu hộ và bảo trì</li> </ul>	<ul style="list-style-type: none"> <li>• Lưu lượng khai thác hạn chế</li> </ul>

### Bơm trục vít (progressive cavity pump)

Ưu điểm	Hạn chế
<ul style="list-style-type: none"><li>• Hiệu suất khá cao</li><li>• Chi phí đầu tư và bảo trì thấp</li><li>• Ít tiêu tốn năng lượng</li><li>• Có khả năng bơm dầu nặng</li><li>• Dễ dàng triển khai lắp đặt</li></ul>	<ul style="list-style-type: none"><li>• Bị giới hạn bởi độ sâu giếng</li></ul>

### Bơm thủy lực (subsurface hydraulic pump)

Ưu điểm	Hạn chế
<ul style="list-style-type: none"><li>• Linh hoạt trong khai thác đa tầng</li><li>• Có thể cho lưu lượng khai thác lớn</li><li>• Tần suất bảo trì thấp</li><li>• Độ sâu khai thác lớn (bơm piston)</li></ul>	<ul style="list-style-type: none"><li>• Chi phí đầu tư ban đầu cao</li><li>• Vận hành phức tạp</li><li>• Hiệu quả kinh tế cao chỉ khi khai thác đa tầng, đa giếng</li><li>• Quá trình bảo trì khó khăn</li></ul>

## 1.2. Trí tuệ nhân tạo

Trí tuệ nhân tạo hay Artificial Intelligence (AI) là một phương pháp sử dụng sức mạnh của máy tính để giải những bài toán khó trong đó có bài toán tối ưu hóa. Để có thể giải được những bài toán tối ưu hóa hiện đại, những kỹ thuật cổ điển không còn thích hợp; tuy nhiên, khi không gian tối ưu lớn đi kèm các ràng buộc đặc biệt cần phải sử dụng những kỹ thuật thích hợp. *Giải thuật di truyền* và *mạng thần kinh nhân tạo* (hay mạng neural nhân tạo, ANN) là hai trong những kỹ thuật đó.

### 1.2.1. Giải thuật di truyền

#### 1.2.1.1. Nguyên lý giải thuật

Giải thuật di truyền (Genetic Algorithms, GA); thuật ngữ “giải thuật” ở đây chính là các thuật toán để giải quyết một vấn đề; được mô phỏng dựa trên các hiện tượng sinh học tự

nhiên: *kế thừa* và *đấu tranh sinh tồn* để thực hiện cải tiến, khảo sát không gian của lời giải. Khái niệm kế thừa và đấu tranh sinh tồn có thể được giải thích như sau:

Giả sử có một quần thể các cá thể loài chồn, bao gồm những cá thể vượt trội có sự linh hoạt, nhạy bén cao, khỏe mạnh và những cá thể yếu ớt không đủ linh hoạt nhạy bén. Những cá thể vượt trội có xác suất sinh tồn trong môi trường tự nhiên cao hơn các cá thể yếu ớt, dĩ nhiên một số cá thể yếu ớt cũng có thể sống được nhờ may mắn đó chính là đấu tranh sinh tồn. Đến một giai đoạn nào đó các cá thể chồn còn lại bắt đầu sinh sản, thế hệ các chú chồn tiếp theo sẽ được kế thừa các đặc tính di truyền của các con bố và mẹ, tất nhiên là có một số con mang đặc tính của những con yếu ớt còn sót lại; điều này chính là kế thừa. Đôi khi sẽ có thêm những cá thể đột biến do mang đặc tính di truyền từ một cá thể khác lạc vào quần thể chồn này, cá thể đột biến này có thể tốt hoặc xấu. Do đó, thế hệ sau luôn mang một xác suất cao cho ra những cá thể tốt hơn thế hệ trước. Tất nhiên quần thể những loài thú săn chồn cũng mang đặc tính kế thừa và đấu tranh này.

Quá trình tìm kiếm nghiệm của bài toán tối ưu bằng giải thuật di truyền cũng thực hiện các bước tương tự với quá trình đấu tranh sinh tồn của quần thể chồn.

Giải thuật di truyền sử dụng một số thuật ngữ được lấy từ di truyền học như: cá thể, quần thể, nhiễm sắc thể, kiểu gen. Một cá thể cũng là một nhiễm sắc thể được cấu tạo từ các gen, mỗi gen mang một đặc trưng khác nhau. Trong toán học một nghiệm của phương trình chính là nhiễm sắc thể, một thành phần trong nghiệm chính là một gen, tập hợp của một số các cặp số có thể trở thành nghiệm của phương trình chính là quần thể. Ví dụ tìm cực trị của hàm số hai biến  $x, y$  sau:

$$f(x, y) = x^2y + y^2x \quad (1.1)$$

Giả sử  $A = (x_1, y_1)$  chính là một nghiệm của 1.1, như vậy  $A$  sẽ là một nhiễm sắc thể;  $x_1, y_1$  là các gen của nhiễm sắc thể  $A$ ; tập hợp  $A_1 \rightarrow A_N$  là một quần thể chứa các nhiễm sắc thể có thể là nghiệm của 1.1 với số lượng cá thể  $N$  tùy thuộc vào sự lựa chọn của người giải.

### 1.2.1.2. Thuật toán di truyền

Một thuật toán di truyền để giải bất kì một bài toán cụ thể nào phải bao gồm năm điểm sau:

- Cách biểu diễn di truyền cho lời giải của bài toán,
- Các khởi tạo quần thể ban đầu,
- Một hàm lượng giá đóng vai trò là môi trường, đánh giá các lời giải theo mức độ “thích nghi” của chúng,
- Các phép toán di truyền,
- Các tham số khởi tạo (kích thước quần thể, xác suất áp dụng các phép toán di truyền, hoán đổi...).

Cụ thể thuật toán di truyền được diễn giải thông qua các phép sau:

#### a. Phép chọn

Phép chọn là quá trình loại bỏ các cá thể xấu trong quần thể chỉ giữ lại những cá thể tốt dựa trên độ thích nghi của mỗi cá thể. Độ thích nghi là một hàm gán một giá trị thực cho các cá thể trong quần thể. Phép chọn có thể được mô phỏng như sau:

- Đánh giá độ thích nghi của từng cá thể trong quần thể hiện hành,
- Sắp xếp quần thể theo độ thích nghi giảm dần,
- Loại bỏ các cá thể cuối dãy, chỉ giữ lại  $n$  cá thể tốt nhất.

#### b. Phép lai

Phép lai là quá trình hình thành nhiễm sắc thể mới trên cơ sở các nhiễm sắc thể bố-mẹ, bằng cách ghép một hay nhiều đoạn gen của hai (hay nhiều) nhiễm sắc thể bố-mẹ với nhau. Phép lai xảy ra với xác suất  $p_c$ , có thể mô phỏng như sau:

- Chọn ngẫu nhiên hai (hay nhiều) cá thể bất kỳ trong quần thể. Giả sử các nhiễm sắc thể bố-mẹ đều có  $m$  gen,



- Tạo ngẫu nhiên một số tự nhiên trong đoạn  $[1, m - 1]$  (điểm lai). Điểm lai chia các chuỗi bố-mẹ dài  $m$  thành hai nhóm chuỗi con dài  $m_1, m_2$ . Hai chuỗi nhiễm sắc thể con mới sẽ là  $m_{11} + m_{22}$  và  $m_{12} + m_{21}$ .

#### c. Phép đột biến

Đột biến là hiện tượng cá thể con mang một số tính trạng không có trong mã gen di truyền của bố-mẹ. Phép đột biến xảy ra với xác suất  $p_m$  nhỏ hơn rất nhiều so với xác suất lai  $p_c$ . Phép đột biến có thể mô phỏng như sau:

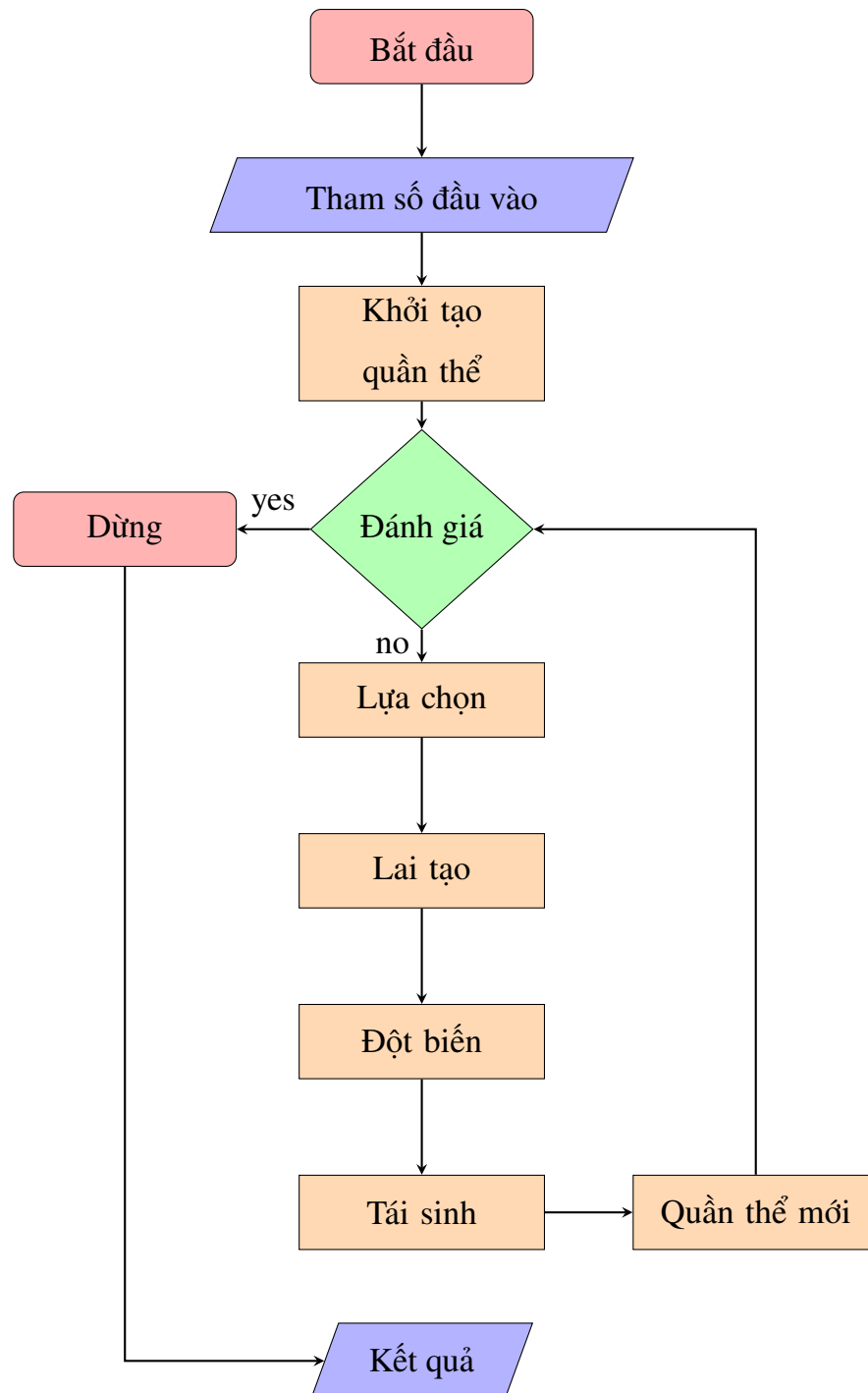
- Chọn ngẫu nhiên một cá thể bất kì bố-mẹ trong quần thể,
- Tạo ngẫu nhiên một số tự nhiên  $k$  trong đoạn  $[1, m]$ ,
- Thay đổi gen thứ  $k$  và trả cá thể này về quần thể để tham gia quá trình tiến hóa tiếp theo.

#### d. Phép tái sinh

Phép tái sinh là quá trình trong đó các cá thể được sao chép trên cơ sở độ thích nghi của nó. Quá trình này có thể được mô tả như sau:

- Tính độ thích nghi của từng cá thể trong quần thể hiện hành, lập bảng cộng dồn các giá trị thích nghi (theo số thứ tự cho từng cá thể). Giả sử quần thể có  $n$  cá thể. Gọi độ thích nghi của cá thể thứ  $i$  là  $F_i$ , tổng cộng dồn thứ  $i$  là  $F_{ti}$ , tổng độ thích nghi của toàn quần thể là  $F_m$ ,
- Tạo một số ngẫu nhiên  $F$  trong đoạn  $[0, F_m]$ ,
- Chọn cá thể thứ  $k$  đầu tiên thỏa  $F \geq F_{tk}$  đưa vào quần thể của thế hệ mới.

Sau khi đã thực hiện phép tái sinh, sẽ bắt đầu một vòng lặp mới cho thế hệ mới sinh từ phép chọn cho đến nghiệm hội tụ hoặc chạy hết số vòng lặp được thiết lập. Thuật toán được thể hiện cụ thể theo sơ đồ Hình 1.4. Điều kiện thực hiện “yes, no” là nghiệm đã hội tụ hoặc đã chạy hết vòng lặp hay chưa. Tham số đầu vào bao gồm xác suất lai tạo ( $p_c$ ) và đột biến ( $p_m$ ), kích thước quần thể ( $N$ ), thành phần nghiệm của phương trình ( $n_{vars}$ ), hàm lượng giá.



Hình 1.4. Thuật toán di truyền

### 1.2.2. Bài toán thời gian

Bài toán thời gian hay chuỗi thời gian có thể coi là quá trình ghi chép lại sự thay đổi giá trị của một đại lượng nào đó (áp suất, nhiệt độ, ...) trong một khoảng thời gian liên tục hoặc gần như liên tục.

Tùy thuộc vào từng loại vấn đề khác nhau mà chuỗi thời gian sẽ mang tầng số khác nhau, có thể là theo từng năm (bài toán ngân sách), quý (bài toán tiêu dùng), tháng (bài toán giao thông), tuần (bài toán doanh số), ngày (bài toán thời tiết), giờ (bài toán cổ phiếu), phút (bài toán cuộc gọi), giây (bài toán mạng) hoặc những bài toán mang tính chất riêng biệt khác.

Mục đích của bài toán thời gian chính là dự báo các giá trị của chuỗi trong tương lai để có thể đem lại lợi ích tốt nhất cho công việc. Dự báo chuỗi thời gian có thể được chia làm hai loại:

- Dự báo chuỗi đơn biến và
- Dự báo chuỗi đa biến.

#### *Tính dừng của chuỗi thời gian*

Để giải quyết bài toán thời gian, một giả định được đưa ra chính là chuỗi dữ liệu đó phải có tính “dừng”, nếu không có giả định này kết quả dự báo sẽ không còn đáng tin cậy. Một chuỗi thời gian được xem là có tính dừng nếu như trung bình, phương sai không đổi theo thời gian và giá trị hiệp phương sai giữa hai đoạn chỉ phụ thuộc vào khoảng cách giữa độ trễ thời gian giữa hai đoạn này chứ không phụ thuộc vào thời điểm thực tế mà hiệp phương sai được tính.

Cụ thể, chuỗi  $Y_t$  được gọi là “dừng” nếu như đồng thời thỏa mãn ba điều kiện:

$$\mu = const \quad \& \quad \sigma^2 = const \quad (1.2)$$

$$Cov(Y_t)_k = E[(Y_t - \mu), (Y_{t+k} - \mu)] \quad (1.3)$$

Trong đó  $\mu, \sigma^2, Cov(Y_t)$  lần lượt là trung bình, phương sai và hiệp phương sai của chuỗi  $Y_t$ .

$Cov(Y_t)_k$  là hiệp phương sai tại độ trễ  $k$  – là phương sai giữa các giá trị  $Y_t, Y_{t+k}$ , tức là hai giá trị của  $Y$  ở các đoạn thời gian cách quãng  $k$ . Nếu  $k = 0$ ,  $Cov(Y_t)_0 = \sigma^2$ ,  $k = 1$   $Cov(Y_t)_1$  là hiệp phương sai giữa hai giá trị kế cận nhau của  $Y$ . Như vậy hiệp phương sai của  $Y$  chỉ phụ thuộc vào  $k$  chứ không phụ thuộc vào thời gian  $t$ . Một chuỗi thời gian có tính dừng sẽ không có xu thế, không có chu kỳ, mà chỉ dao động xung quanh giá trị kỳ vọng của nó.

Trong thực tế, phần lớn các dữ liệu chuỗi thời gian đều không dừng, nếu như muốn thực hiện dự báo trên các chuỗi này cần có những phương pháp để khắc phục tính không dừng của chuỗi. Một trong các phương pháp đó là sử dụng toán tử sai phân lùi.

Để có thể dễ hiểu hơn, phương pháp cơ bản được triển khai như sau:

$$\mathbb{L}(Y_t) = Y_{t-1} \quad (1.4)$$

Với  $\mathbb{L}$  chính là một lần thực hiện toán tử lùi; nói cách khác, thực hiện  $\mathbb{L}$  trên  $Y_t$ , có tác dụng dịch chuyển dữ liệu trở lại một đoạn thời gian.

Áp dụng  $\mathbb{L}$  hai lần trên  $Y_t$  sẽ chuyển dữ liệu trở lại hai đoạn thời gian:

$$\mathbb{L}(\mathbb{L}(Y_t)) = \mathbb{L}^2(Y_t) = Y_{t-2} \quad (1.5)$$

Toán tử lùi giúp quá trình tính sai phân được dễ dàng hơn. Với sai phân bậc nhất:

$$\mathbb{D}(Y_t) = Y_t - Y_{t-1} \quad (1.6)$$

Với  $\mathbb{D}$  là phép tính sai phân. Kết hợp với toán tử lùi, có thể viết lại như sau:

$$\mathbb{D}(Y_t) = Y_t - \mathbb{L}(Y_t) = (1 - \mathbb{L})Y_t \quad (1.7)$$

Tương tự, với sai phân bậc hai sẽ có:

$$\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1} = 1 - 2\mathbb{L} + \mathbb{L}^2 Y_t = (1 - \mathbb{L})^2 Y_t \quad (1.8)$$

Tổng quát nếu lấy sai phân bậc  $d$  sẽ đạt được tính dừng của chuỗi thời gian:

$$\Delta^d Y_t = (1 - \mathbb{L})^d Y_t \quad (1.9)$$

### 1.2.3. Mạng neuron truyền thẳng đa lớp

Mạng neuron truyền thẳng đa lớp (Multilayer Perceptron, MLP hoặc feedforward network) là một kỹ thuật trí tuệ nhân tạo, được xem như là một trong những tinh túy của kỹ thuật học sâu. Mục đích khi xây dựng kỹ thuật này chính là tạo ra một hàm số xấp xỉ  $f$  có khả năng miêu tả được mối quan hệ giữa hai đại lượng đầu vào ( $\mathbf{x}$ ) và đầu ra ( $\mathbf{y}$ ), đơn giản là  $\mathbf{y} = f(\mathbf{x}; \theta)$  với  $\theta$  là tập hợp các tham số đạt được trong quá trình tính toán. Mạng truyền thẳng là một kỹ thuật cực kì quan trọng trong các kỹ thuật trí tuệ nhân tạo, là nền tảng của hầu hết các kỹ thuật tiên tiến khác như mạng tích chập hay mạng hồi quy. Đây đều là những mạng cực kì phức tạp và có ý nghĩa to lớn trong ngành công nghiệp hiện đại.

Lấy cảm hứng từ hệ neuron thần kinh sinh học, với tính chất truyền thông tin thẳng thông qua hàm số từ các giá trị đầu vào tới các phép tính trung gian xác định  $f$  đến điểm cuối là các giá trị đầu ra; những mô hình xác định hàm  $f$  theo cách đó được gọi là mô hình truyền thẳng; hàm  $f$  lúc này bao gồm các thành phần hàm số khác nhau đại diện cho từng vị trí liên kết tạo thành một cấu trúc chuỗi, cấu trúc chuỗi này tạo nên mạng neuron truyền thẳng hay ngắn gọn là mạng truyền thẳng. Ví dụ, một chuỗi bao gồm ba hàm  $f_1, f_2$  và  $f_3$  sẽ có cấu trúc như sau:

$$f(\mathbf{x}) = f_3(f_2(f_1(\mathbf{x}))) \quad (1.10)$$

$f(\mathbf{x})$  chính là hàm xấp xỉ cần tìm, các hàm  $f_1, f_2, f_3$  là các hàm liên kết. Trong trường hợp này hàm  $f_1$  đại diện cho lớp thứ nhất của mạng, hàm  $f_2$  đại diện cho lớp thứ hai và cứ tiếp tục như vậy cho tới lớp thứ  $n$ . Chiều dài của chuỗi hay “độ sâu” của mô hình sẽ được xem như là một trong các nhân tố tạo nên mức độ phức tạp của mạng. Phân lớp cuối cùng của cấu trúc được gọi là phân lớp đầu ra, các phân lớp trước phân lớp đầu ra được gọi là phân lớp ẩn; được gọi là “ẩn” vì các phân lớp này không hiển thị các giá trị kết quả được tính toán trong phân lớp đó mà truyền thẳng sang cho phân lớp tiếp theo. Mỗi phân lớp ẩn của mạng được xem như là một vector, lúc này, chiều của vector hay chiều của phân lớp ẩn chính là “độ rộng” của mô hình, là nhân tố thứ hai tạo nên độ phức tạp của mạng. Mỗi phần tử trong vector đóng vai trò như là một neuron hay *unit*, thông thường mỗi *unit* sẽ

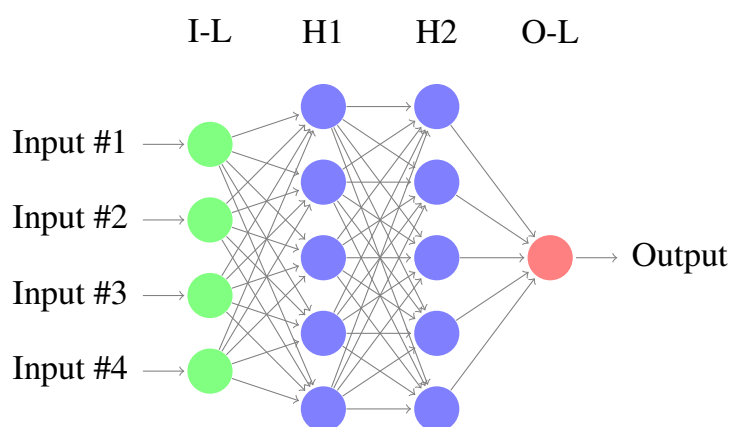
nhận tín hiệu từ các *unit* ở phân lớp trước nó.

Để đưa một mạng truyền thẳng vào sử dụng cần xác định các tham số cơ bản đi kèm như độ sâu, độ rộng, số lượng *unit* trong mỗi phân lớp kèm theo đó là các hàm kích hoạt, hàm tối ưu và phương pháp đánh mức độ chính xác của hàm  $f$ .

### 1.2.3.1. Cấu trúc cơ bản

Xác định cấu trúc cơ bản của mạng cũng chính là xác định các thông số về độ sâu ( $N_l$ ) và độ rộng ( $N_n$ ) của mô hình. Để có được một kết quả tốt,  $N_l$  và  $N_n$  phải phù hợp với bài toán đang giải quyết, quá đơn giản hoặc quá phức tạp đều đem lại một kết quả kém chính xác. Trong thực tế không có một phương pháp cụ thể nào để xác định các thông số này, hầu hết  $N_l$  và  $N_n$  đều được xác định dựa trên cảm quan của người giải quyết bài toán đối với dữ liệu đầu vào của bài toán, điều này thực sự rất khó khăn đối với những người mới bắt đầu. Do đó, một vài cá nhân đã đưa ra một cách không chính quy đó là dựa vào số lượng các thuộc tính của dữ liệu sau khi xử lý để xác định  $N_l$  và  $N_n$ . Mặc dù không chính quy, tuy nhiên kết quả đem lại từ cách này cũng khá khả quan đối với các bài toán đơn giản có dữ liệu đầu vào thuộc dạng dữ liệu cấu trúc (có thể diễn tả dưới dạng số).

Về cơ bản, một mạng truyền thẳng có thể được thể hiện như sau:



Hình 1.5. Cấu trúc của một mạng truyền thẳng

Với Input, Output, I-L, O-L lần lượt là các dữ liệu đưa vào mạng, kết quả đạt được, phân

lớp đầu vào, phân lớp đầu ra và các phân lớp ẩn H1, H2. Có thể thấy được mạng này có độ sâu là 3 (phân lớp đầu vào không được tính vào độ sâu của mạng), độ rộng là 5 với 2 phân lớp ẩn. Một mạng như vậy cũng được gọi là mạng liên kết toàn phần.

#### 1.2.3.2. Hàm kích hoạt

Về cơ bản hàm kích hoạt chính là để chuẩn hóa các thông tin đầu ra của mỗi unit. Phức tạp hơn, nếu coi dữ liệu đầu vào là một ma trận hoặc một vector, khi áp dụng hàm kích hoạt cho ma trận (hoặc vector), nó sẽ được áp dụng cho từng thành phần của ma trận đó. Sau đó sắp xếp lại đúng theo thứ tự để được một ma trận có kích thước bằng với ma trận đầu vào của phân lớp tiếp theo.

Một vài hàm kích hoạt nổi tiếng như *Linear*, *Sigmoid*, *Softmax*, *Tanh*, *ReLU* ngoài ra còn nhiều biến thể khác tuy nhiên sẽ không được nhắc tới trong phạm vi của luận án.

##### a. Hàm Linear

Hàm *Linear* thuộc loại hàm kích hoạt đơn giản nhất, biến đổi các tham số đầu vào theo hàm tuyến tính  $y = w.x + b$ . Nhược điểm của hàm *Linear* là không thể biến đổi phi tuyến, thường được sử dụng trong các bài toán đơn giản.

##### b. Hàm Sigmoid

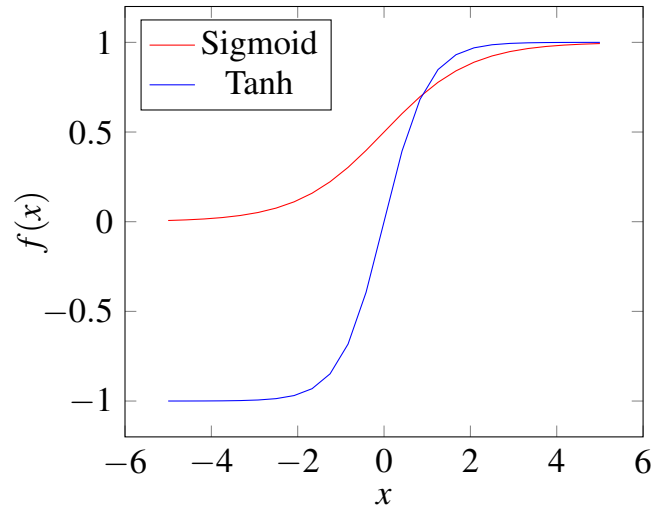
Hàm *Sigmoid* biến đổi tham số đầu vào theo dạng:

$$y = \frac{1}{1 + e^{-x}} \quad (1.11)$$

Có thể thấy được đầu ra của hàm *Sigmoid* nằm trong khoảng  $[0, 1]$  (Hình 1.6), do đó hàm *Sigmoid* thường được kết hợp với một số kỹ thuật khác để áp dụng cho bài toán phân loại nhị phân. Hiện tại, hàm *Sigmoid* không còn được sử dụng nhiều cho các bài toán phức tạp do hiện tượng tiêu biến giá trị tối ưu trong quá trình huấn luyện của mạng.

##### c. Hàm Tanh

Hàm *Tanh* khá tương tự hàm *Sigmoid* với giá trị đầu ra nằm trong khoảng  $[-1, 1]$  (Hình



Hình 1.6. Hàm *Tanh* và *Sigmoid*

1.6), đồng thời hàm *Tanh* cũng mang nhược điểm chính đó là làm tiêu biến giá trị tối ưu. Tuy nhiên, vẫn có một số biến thể của hàm *Tanh* còn được sử dụng. Hàm *Tanh* có dạng như sau:

$$y = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (1.12)$$

#### d. Hàm softmax

Hàm *Softmax* là một hàm rất đặc biệt có thể biến đổi đầu vào dạng số trở thành đầu ra dạng xác suất có tổng bằng một. Hàm *Softmax* cực kì thích hợp cho các bài toán phân loại. Hàm *Softmax* có dạng như sau:

$$y_i = \frac{e^{x_i}}{\sum e^{x_i}} \quad (1.13)$$

Có thể lấy một ví dụ như sau đối với hàm *Softmax*, các giá trị đầu ra sau biến đổi luôn có tổng bằng một.

$$x_i = \begin{bmatrix} 1.5 \\ 0.7 \\ 0.4 \end{bmatrix} \Rightarrow y_i = \frac{e^{x_i}}{\sum e^{x_i}} \Rightarrow \begin{bmatrix} p = 0.58 \\ p = 0.27 \\ p = 0.15 \end{bmatrix} \quad (1.14)$$

#### e. Hàm ReLU

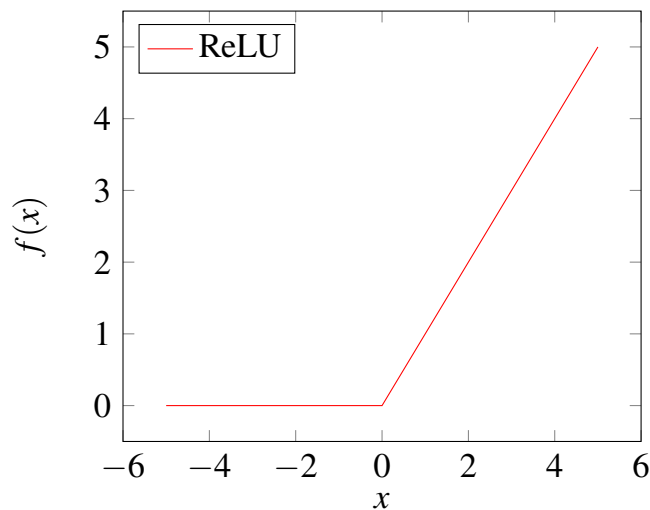


Hàm *ReLU* (Rectified Linear Unit) và các biến thể của nó được sử dụng phổ biến nhất trong thời điểm hiện tại. Ưu điểm lớn nhất của nó chính là loại bỏ được hiện tượng tiêu biến giá trị tối ưu xuất hiện ở hàm *Tanh*, *Sigmoid*. Hàm *ReLU* biến đổi các giá trị đầu vào theo dạng hàm số rất đơn giản:

$$y = \max(0, x) \quad (1.15)$$

Kết hợp đồ thị Hình 1.7, hàm số trên có thể hiểu như sau:

$$y = \begin{cases} 0, & \text{khi } x < 0 \\ x, & \text{khi } x \geq 0 \end{cases} \quad (1.16)$$



Hình 1.7. Hàm ReLU

### 1.2.3.3. Hàm tối ưu

Trong quá trình tính toán một hàm số được đưa ra để thể hiện sai số giữa các giá trị dự đoán và giá trị thực tế, hàm số này được gọi là hàm mất mát. Thực tế quá trình tính toán chính là tìm điểm tối ưu của hàm này, tức là tìm điểm nhỏ nhất của nó. Đó chính là nhiệm vụ của hàm tối ưu. Hàm tối ưu bao gồm các thuật toán toán học nhằm mục đích đưa giá trị hàm mất mát xuống mức thấp nhất. Giá trị sai số này càng thấp, kết quả đem lại càng tốt.

Một vài thuật toán nổi tiếng có thể kể đến như *Gradient Descent*, *Adam*, *RMSProp*, *AdaDelta*, *AdaGrad*. Về cơ bản các thuật toán này đều có chung một quy trình tính toán như sau:

1. Tính toán sự ảnh hưởng khi thay đổi trọng số đối với hàm mất mát.
2. Thay đổi trọng số dựa trên sự ảnh hưởng đó.
3. Lặp lại bước 1 và 2 cho đến khi hàm mất mát đạt giá trị nhỏ nhất có thể.

Trong đó có thể nói thuật toán *Adam* và các biến thể của nó đang được sử dụng phổ biến nhất thời điểm hiện tại do khả năng hội tụ nghiệm nhanh chóng.

#### 1.2.3.4. Phương pháp đánh giá

Phương pháp đánh giá (metrics) được sử dụng để tính toán định lượng chất lượng của kết quả dự đoán so với các giá trị thực tế, đồng thời cho người dùng một cái nhìn khách quan về chất lượng của mô hình. Các phương pháp đánh giá như *căn bậc hai trung bình bình phương sai số (RMSE)*, *trung bình sai số tuyệt đối (MAE)* thường được sử dụng cho một bài toán hồi quy. Tuy nhiên một bài toán thời gian cần các phương pháp đánh giá đặc biệt hơn thích hợp hơn cho một chuỗi thời gian liên tục: *trung bình phần trăm sai số tuyệt đối (MAPE)*, *trung bình tỉ lệ sai số tuyệt đối (MASE)*. Các phương pháp này được tính toán như sau:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=0}^N (y_i - \hat{y}_i)^2} \quad (1.17)$$

$$MAE = \frac{1}{N} \sum_{i=0}^N |y_i - \hat{y}_i| \quad (1.18)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{|y_i|} \quad (1.19)$$

$$MASE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{\frac{N}{N-1} \sum_{i=2}^N |y_i - y_{i-1}|} \quad (1.20)$$

Trong đó  $y_i$  là giá trị thực,  $\hat{y}_i$  là giá trị dự đoán,  $N$  là tổng số dữ liệu. Trong quá trình đánh giá, giá trị tính toán từ các phương pháp này càng thấp càng tốt.

### 1.3. Đường cong đặc tính gas lift

Mục tiếp chính khi sử dụng phương pháp gas lift chính là thu được lưu lượng dầu tối đa với lưu lượng gas lift tối ưu nhất. Mối quan hệ giữa hai đại lượng này được thể hiện thông

qua đường cong đặc tính gas lift hay Gas Lift Performance Curve (GLPC). Lưu lượng khí bơm ép tối ưu rất quan trọng, bơm dư hoặc thiếu khí đều làm giảm lưu lượng khai thác đồng thời tăng chi phí bơm ép. Mô hình phân phối lưu lượng khí sẽ được xây dựng dựa trên các tính chất của GLPC cho một giếng khai thác. Dòng chảy trong vỉa được giải sử chỉ bao gồm một pha lỏng, trong tubing sẽ là hai pha lỏng - khí, các điều kiện vật lý để tồn tại GLPC là duy nhất.

### **1.3.1. Đường cong đặc tính và bài toán tối ưu**

Đối với một giếng khai thác bắt đầu thực hiện gas lift, sản phẩm khai thác sẽ bao gồm các lưu chất vỉa (dầu, nước) và khí được bơm ép vào trong tubing. Khi đó lưu lượng khai thác có thể được biểu diễn bằng một hàm phụ thuộc vào lưu lượng khí bơm ép:

$$q_l = \varphi(q_g) \quad (1.21)$$

Với:

$q_g$  là lưu lượng khí bơm ép

$q_l$  là lưu lượng khai thác.

Nếu hàm số  $\varphi(q_g)$  tồn tại, đồ thị biểu diễn đường cong hàm số 1.21 chính là đường đặc tính gas lift (Hình 1.8). Để 1.21 tồn tại  $q_g$  phải thỏa mãn điều kiện  $0 \leq q_g \leq Q_{gav}$  với  $Q_{gav}$  là tổng lượng khí có sẵn.

Sử dụng các đặc tính khai thác của mỗi giếng lưu lượng dầu được tính thông qua phương trình:

$$q_o = (1 - WC) \times q_l \quad (1.22)$$

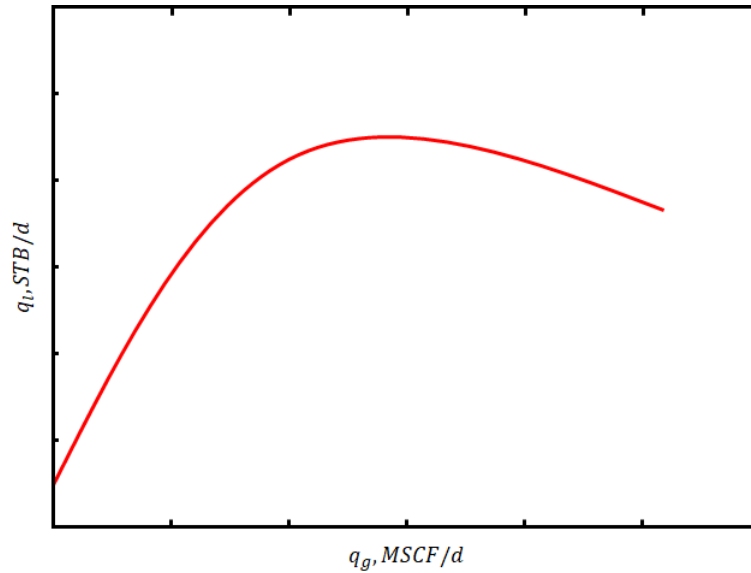
Với:

WC là hàm lượng nước trong lưu chất khai thác

$q_o$  là lưu lượng dầu.

Từ các phương trình 1.21 và 1.22 được:

$$q_o = (1 - WC) \times \varphi(q_g) \quad (1.23)$$



Hình 1.8. Đường đặc tính gas lift

Quá trình tối đa hóa lưu lượng dầu sẽ xoay quanh việc tìm giá trị lớn nhất của phương trình 1.23:

$$\text{Max} \sum_{k=1}^N q_o = \text{Max} \sum_{k=1}^N (1 - WC) \times \varphi(q_g) \quad (1.24)$$

Với N là tổng số giếng trong cụm giếng và k là giếng khai thác.

### 1.3.2. Mô hình toán học

Để có thể thực hiện xây dựng mô hình phân phối tối ưu cần đưa bài toán tối ưu như đã nêu ở mục 1.3.1 trở về dưới dạng tối ưu bằng mô hình toán học. Bắt đầu với phương trình biểu diễn chỉ số năng suất khai thác được suy ra từ định luật Darcy:

$$J = \frac{q_l}{P_r - P_{wf}} \quad (1.25)$$

Với:

J là chỉ số năng suất

$P_r$  là áp suất vỉa

$P_{wf}$  là áp suất dòng vào tại đáy giếng.

Bằng cách giả sử lưu chất vỉa chỉ tồn tại một pha lỏng (dầu và nước), ta có được:

$$q_l = (1 + WOR) \times q_o = \frac{q_o}{1 - WC} \quad (1.26)$$

Với WOR là tỉ số nước - dầu.

Phương trình dòng chảy cho lưu chất hai pha trong tubing được biểu diễn từ *phương trình cân bằng năng lượng*:

$$\frac{dP}{dz} = \frac{g}{g_c} \bar{\rho} \sin(\theta) + \frac{f \bar{\rho} u_m^2}{2g_c D} + \frac{\bar{\rho}}{2g_c} \frac{D}{dz} u_m^2 \quad (1.27)$$

Với điều kiện ban đầu:

$$P(0) = P_{wh} \quad (1.28)$$

Trong đó:

P: Áp suất mất mát (psi)

z: Chiều sâu giếng khoan (ft)

g: Gia tốc trọng trường (ft/s<sup>2</sup>)

g<sub>c</sub>: Hệ số chuyển đổi

$\bar{\rho}$ : Tỉ trọng hỗn hợp

D: Đường kính tubing (in)

f: Hệ số ma sát

u<sub>m</sub>: Vận tốc hỗn hợp (ft/s).

Các thành phần  $\frac{g}{g_c} \bar{\rho} \sin(\theta)$ ,  $\frac{f \bar{\rho} u_m^2}{2g_c D}$  và  $\frac{\bar{\rho}}{2g_c} \frac{d}{dz} u_m^2$  trong phương trình 1.27 tương ứng là tổn hao áp suất do thế năng, ma sát và động năng. Đối với giếng thẳng đứng  $\sin(\theta) = 1$  do đó thành phần thế năng trở thành  $\frac{g}{g_c} \bar{\rho}$ , thành phần động năng tương đối nhỏ có thể bỏ qua.  $\bar{\rho}$  được miêu tả bằng phương trình:

$$\bar{\rho} = H_g \rho_g + (1 - H_g) \rho_l \quad (1.29)$$

Tỉ trọng của khí  $\rho_g$  phụ thuộc áp suất P theo công thức:

$$\rho_g = 28.97 \frac{\gamma_g P}{ZRT} \quad (1.30)$$

Với:

Z: Hệ số lệch khí

R: Hằng số khí phổ

T: Nhiệt độ trung bình trong tubing.

Tỉ số thể tích khí trong hỗn hợp  $H_g$  được tính theo tương quan Zuber-Findlay [cite]:

$$H_g = \frac{u_{sg}}{C_o u_m + U_d} \quad (1.31)$$

Với:

$u_{sg}$ : Tốc độ trượt của khí

$C_o$ : Hệ số phân tán

$U_d$ : Tốc độ trôi.

Dựa vào 1.31 có thể tính được  $H_g$  khi biết được các hệ số phân tán  $C_o$  và tốc độ trôi  $U_d$ .  $C_o$  và  $U_d$  cũng được Zuber-Findlay tính toán ra dựa trên thực nghiệm, trong đó  $C_o = 1.2$  và  $U_d$  được tính theo công thức:

$$U_d = 1.53 \left( \frac{g \sigma \Delta \rho}{\rho_L^2} \right)^{0.25} \quad (1.32)$$

Giá trị  $\sigma$  chính là sức căng bề mặt của lưu chất được tính theo các thông số PVT.

Tốc độ của hỗn hợp lưu chất được tính toán từ tốc độ trượt của pha khí và pha lỏng theo các công thức:

$$u_m = u_{sl} + u_{sg} \quad (1.33)$$

$$u_{sl} = \frac{q_l}{A}, \quad u_{sg} = \frac{Z P_{sc} T}{T_{sc} P} \frac{q_g}{A}, \quad A = \pi \frac{D^2}{4} \quad (1.34)$$

Với:

$u_{sl}$ : Tốc độ trượt của pha lỏng (ft/s<sup>2</sup>)

$u_{sg}$ : Tốc độ trượt của pha khí (ft/s<sup>2</sup>)

A: Tiết diện của tubing.

Hệ số ma sát  $f$  trong 1.27 được tính toán từ hệ số Reynolds  $Re$  với các chế độ dòng chảy,

độ nhám của tubing khác nhau bằng phương trình Colebrook [cite]:

$$\frac{1}{\sqrt{f}} = 2 \times \log_{10} \left( \frac{\varepsilon/D_h}{3.7} + \frac{2.51}{Re\sqrt{f}} \right) \quad (1.35)$$

Trong đó:

$\varepsilon$ : độ nhám của tubing

$D_h$ : đường kính thủy lực

$Re$ : Hệ số Reynolds.

Với tubing có tiết diện tròn đường kính thủy lực  $D_h$  chính là đường kính của tubing.

Để có thể thực hiện tính toán một cách dễ dàng hơn, các thông số như áp suất, độ sâu, lưu lượng, tỉ trọng và vận tốc sẽ được tiêu chuẩn hóa bằng cách sử dụng những công thức chuẩn hóa sau:

$$\tilde{P} = \frac{P}{P_r}, \quad \tilde{z} = \frac{z}{L} \quad (1.36)$$

$$\tilde{q}_g = \frac{q_g}{LD\sqrt{gD}}, \quad \tilde{q}_l = \frac{q_l}{JP_r} \quad (1.37)$$

$$\tilde{\rho} = \frac{\bar{\rho}Lg}{g_cP_r}, \quad \tilde{u} = \frac{u_m}{\sqrt{gD}} \quad (1.38)$$

Như vậy công thức 1.27 và 1.25 sẽ trở thành:

$$\frac{d\tilde{P}}{d\tilde{z}} = \tilde{\rho} \left( 1 + \frac{f\tilde{u}^2}{2} \right) \quad (1.39)$$

$$\tilde{P}_{wf} = 1 - \tilde{q}_l \quad (1.40)$$

Do sự thay đổi của  $q_l$  và  $q_g$  theo từng giếng nên phương trình đạo hàm tổn hao áp suất có thể được đưa về dạng:

$$\frac{dP_k}{dz_k} = P_k(z_k; q_{gk}, q_{lk}) \quad (1.41)$$

Như vậy, có thể suy ra một phương trình tổng quát từ 1.39 và 1.41 như sau:

$$\tilde{P}_k(\tilde{z}_k; \tilde{q}_{gk}, \tilde{q}_{lk}) = \tilde{\rho}_k \left( 1 + \frac{f\tilde{u}_k^2}{2} \right) \quad (1.42)$$

Khi  $\tilde{z}_k = 0$ , áp suất  $P_k(0)$  là điều kiện ban đầu của phương trình 1.42; khi  $\tilde{z}_k = 1$ , áp suất  $P_k$  chính là áp suất tại đáy giếng.

$$\tilde{P}_k(0; \tilde{q}_{gk}, \tilde{q}_{lk}) = \tilde{P}_{whk} \quad (1.43)$$

$$\tilde{P}_k(1; \tilde{q}_{gk}, \tilde{q}_{lk}) = 1 - \tilde{q}_{lk} \quad (1.44)$$

Với  $k = 1, 2, \dots, N$ ,  $N$  là tổng số giếng trong hệ thống cụm giếng và đặc tính gas lift của giếng thứ  $k$  là:

$$q_{lk} = \varphi_k(q_{gk}) \quad (1.45)$$

Quay lại với bài toán tối ưu đã nêu ở mục 1.3.1, vấn đề cần giải quyết ở đây chính là tối đa lưu lượng dầu khai thác được trong hệ thống cụm giếng thực hiện bơm ép gas lift. Phương trình 1.23 được miêu tả cụ thể như sau:

$$\max(Q_o) = \sum_{k=1}^N (1 - WC_k) \times \varphi_k(q_{gk}) \quad (1.46)$$

Với điều kiện:

$$\sum_{k=1}^N q_{gk} \leq Q_{gav} \quad (1.47)$$

Bài toán tìm kiếm nghiệm để phương trình 1.46 đạt giá trị lớn nhất có thể được đưa về trở thành bài toán tìm nghiệm để 1.48 đạt giá trị nhỏ nhất:

$$\min \Omega(\bar{q}_g, \bar{q}_l) = \frac{1}{\sum_{k=1}^N (1 - WC_k) q_{lk}} \quad (1.48)$$

Trong không gian tìm kiếm:

$$D = \left\{ (\bar{q}_g, \bar{q}_l) \in \mathbb{R}^{2N} \mid \sum_{k=1}^N q_{gk} \leq Q_{gav} \right\} \quad (1.49)$$

### 1.3.3. Giải pháp số

Để giải quyết bài toán cực tiểu hóa với các ràng buộc như đã nêu ở trên, một phương pháp được đưa ra đó chính là kết hợp giải thuật di truyền với hàm phạt (penalty). Hàm phạt sẽ hướng các cá thể vào không gian tìm kiếm hợp lệ bằng cách cộng thêm các giá trị phạt nếu



như một cá thể không phù hợp với điều kiện ràng buộc của bài toán. Khi các cá thể ở trong miền hợp lệ thì giá trị phạt bằng 0, nếu cá thể ở càng xa miền hợp lệ giá trị phạt tăng lên.

Để sử dụng phương pháp này cần xác định giá trị của hàm phạt, thông qua hệ số phạt  $\lambda$ , một cách phù hợp. Việc xác định giá trị hàm phạt tùy thuộc vào từng vấn đề tối ưu hóa cụ thể. Với bài toán cực tiểu hóa phương trình 1.48, hàm phạt được thiết lập dựa trên phương trình đạo hàm tổn hao áp suất 1.42. Điều kiện là áp suất tổn hao tại điểm tính toán không được lớn hơn áp suất dòng vào tại đáy giếng, tức là:

$$\tilde{P}_k(\tilde{z}_k; \tilde{q}_{gk}, \tilde{q}_{lk}) \leq 1 - \tilde{q}_{lk} \quad (1.50)$$

Phương trình 1.48 có thể đưa về dạng kết hợp hàm phạt như sau:

$$\min \hat{\Omega}(\bar{q}_g, \bar{q}_l) = \frac{1}{\sum_{k=1}^N (1 - WC_k) q_{lk}} + \lambda \max\{0, \tilde{P}_k(\tilde{z}_k; \tilde{q}_{gk}, \tilde{q}_{lk}) - (1 - \tilde{q}_{lk})\} \quad (1.51)$$

Trong không gian tìm kiếm:

$$\hat{D} = \left\{ (\bar{q}_g, \bar{q}_l) \in \mathbb{R}^{2N} \mid \sum_{k=1}^N q_{gk} \leq Q_{gav}, 0 \leq q_{lk} \leq 1, k = 1, 2, \dots, N \right\} \quad (1.52)$$

$\lambda$  là hệ số phạt, nghiệm của 1.48 sẽ hội tụ khi  $\lambda$  đủ lớn. Vì 1.42 là một phương trình đạo hàm chưa biết giá trị cụ thể nên sẽ được tính toán thông qua phương pháp xấp xỉ Runge-Kutta bậc 4.

#### 1.4. Thuật toán tối ưu

Thuật toán di truyền kết hợp với hàm phạt sẽ được xây dựng cho phương trình 1.51 trong không gian  $\hat{D}$  như sau:

1. Khởi tạo không gian quần thể *POP* kích thước  $n$  với các nhiễm sắc thể  $v_1, v_2, \dots, v_n$  tương ứng với các cặp  $(\bar{q}_g^{(i)}, \bar{q}_l^{(i)})$ ,  $i = 1, 2, \dots, n \subset \hat{D}$ .
2. Tính toán giá trị  $\tilde{P}_k(\tilde{z}_k; \bar{q}_g^{(i)}, \bar{q}_l^{(i)})$  theo các nhiễm sắc thể  $v_1, v_2, \dots, v_n$  bằng phương pháp Runge-Kutta bậc 4 và đưa ra giá trị hàm phạt.

3. Đánh giá giá trị  $\widehat{\Omega}(\bar{q}_g^{(i)}, \bar{q}_l^{(i)})$ .
4. Tạo các nhiễm sắc thể mới bằng phép lai và phép đột biến.
5. Sử dụng phép chọn để tạo một quần thể mới kích thước không đổi.
6. Bắt đầu lại với bước 2 cho đến khi nghiệm hội tụ hoặc kết thúc vòng lặp.